

What is claimed is:

1. A method of personalizing search results of a search engine, comprising:
 - accessing a user profile for a user based on information about the user, the user information including information derived from a set of documents, the set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from the search engine, documents accessed by the user, documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the user;
 - receiving a search query from the user;
 - identifying a set of search result documents that match the search query;
 - assigning a generic score to each of at least a plurality of the search result documents;
 - assigning a personalized score to each document of the plurality of search result documents in accordance with the generic score assigned to the document and the user profile; and
 - ranking the set of search result documents according to their personalized scores.
2. The method of claim 1, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have been viewed by the user.
3. The method of claim 1, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have not been viewed by the user.
4. The method of claim 1, including updating the user profile by:
 - updating a term-based profile by identifying a set of terms from a document in the set of documents, and adding information about the identified set of terms to the term-based profile; and
 - updating a category-based profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile.
5. The method of claim 4, including updating the user profile by:

updating a link-based profile by analyzing links in the document, and adding information derived from the analyzed links to the link-based profile.

6. The method of claim 5, wherein the link-based profile includes information about URLs or portions of URLs.

7. The method of claim 6, wherein the link-based profile comprises:

a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the user visits the URL, time the user has spent viewing a document associated with the URL and quantity of the user's scrolling activity at the document; and

a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the user's visits to the host.

8. The method of claim 7, wherein the URLs further include URLs that have not been visited by the user, but are related to the URLs that have been visited by the user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

9. The method of claim 4, wherein a term in the term-based profile is an expression comprising at least one word and a weight.

10. The method of claim 9, wherein the weight is a weight associated with occurrences of the term in the set of documents.

11. The method of claim 9, wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said set of documents.

12. The method of claim 4, wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories.

13. The method of claim 12, wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin.

14. The method of claim 1, including updating the user profile by:
updating a category-based profile by classifying a document in the set of documents into a plurality of categories, and adding information about the plurality of categories to the category-based profile.
15. The method of claim 1, wherein the user profile is stored on a server of the search engine.
16. The method of claim 1, wherein the user profile is stored on a client associated with the user.
17. The method of claim 1, wherein the user is a group of users.
18. A method of personalizing search results of a search engine, comprising:
creating a plurality of user profiles for a plurality of users, each user profile including at least a user's identification number and information derived from documents visited by the user;
receiving a search query from a user of the plurality of users, the search query including at least one query term and the user's identification number;
selecting a set of documents from the Internet, assigning to each document in the set a generic score that characterizes the relevance of the document to the at least one query term;
retrieving the user's user profile and assigning to each of the set of documents a profile score based on the user profile; and
ranking the set of documents according to their generic and profile scores.
19. The method of claim 18, wherein the step of creating a user's user profile further comprises:
creating a term-based profile by extracting a set of terms from documents visited by the user and associating a weight with each extracted term; and
creating a category-based profile by determining a plurality of categories associated with documents visited by the user and associating a weight with each determined category.
20. The method of claim 18, wherein the step of creating a user's user profile further comprises:

creating a link-based profile by analyzing links in documents visited by the user and associating weights with the link.

21. The method of claim 18, wherein the user profile for a particular user includes demographic and geographic information provided by the user.

22. The method of claim 18, wherein the documents visited by the user from which information is derived for use in a particular user's user profile are selected based on the user's activities when visiting the documents.

23. The method of claim 18, including storing the plurality of user profiles on a server of the search engine; and

the retrieving including identifying the user's user profile based on a user identification number associated with both the user and the user's user profile.

24. The method of claim 18, including storing the plurality of user profiles on client computers associated with the plurality of users.

25. The method of claim 18, wherein the ranked set of documents comprises two sets of documents, one set of documents ordered by their generic scores and the other set of documents ordered by personalized scores generated by combining the document's generic and profile scores.

26. The method of claim 25, including interleaving or intermixing the two sets of documents to form the ranked set of documents.

27. A search engine system, comprising:

one or more central processing units for executing programs;

an interface for receiving event messages; and

a search engine module executable by the one or more central processing units, the module comprising:

instructions for accessing a user profile for a user, the user profile based on information about the user, the user information including information derived from a set of documents, the set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from the search engine system,

documents accessed by the user, documents linked to the documents identified by search results from the search engine system, and documents linked to the documents accessed by the user;

instructions for receiving a search query from the user;

instructions for identifying a set of search result documents that match the search query;

instructions for assigning a generic score to each of at least a plurality of the search result documents;

instructions for assigning a personalized score to each document of the plurality of search result documents in accordance with the generic score assigned to the document and the user profile; and

instructions for ranking at least the plurality of the search result documents according to their personalized scores.

28. The system of claim 27, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have been viewed by the user.

29. The system of claim 27, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have not been viewed by the user.

30. The system of claim 27, further including:

instructions for updating a term-based profile by identifying a set of terms from a document in the set of documents, and adding information about the identified set of terms to the term-based profile; and

instructions for updating a category-based profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

31. The system of claim 30, further including:

instructions for updating a link-based profile by analyzing links in the document, and adding information derived from the analyzed links to the link-based profile.

32. The system of claim 31, wherein the link-based profile includes information about URLs or portions of URLs.
33. The system of claim 32, wherein the link-based profile comprises:
a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the user visits the URL, time the user has spent viewing a document associated with the URL and quantity of the user's scrolling activity at the document; and
a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the user's visits to the host.
34. The system of claim 33, wherein the URLs further include URLs that have not been visited by the user, but are related to the URLs that have been visited by the user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.
35. The system of claim 30, wherein a term in the term-based profile is an expression comprising at least one word and a weight.
36. The system of claim 35, wherein the weight is a weight associated with occurrences of the term in the set of documents.
37. The system of claim 35, wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said set of documents.
38. The system of claim 30, wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories.
39. The system of claim 38, wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin.
40. The system of claim 27, wherein the user profile is stored on a server of the search engine.

41. The system of claim 27, wherein the user profile is stored on a client associated with the user.

42. The system of claim 27, wherein the user is a group of users.

43. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

instructions for accessing a user profile for a user based on information about the user, the user information including information derived from a set of documents, the set of documents comprising a plurality of documents selected from the set consisting of documents identified by search results from a search engine, documents accessed by the user, documents linked to the documents identified by search results from the search engine, and documents linked to the documents accessed by the user;

instructions for receiving a search query from the user;

instructions for identifying a set of search result documents that match the search query;

instructions for assigning a generic score to each of at least a plurality of the search result documents;

instructions for assigning a personalized score to each document of the plurality of search result documents in accordance with the generic scores assigned to the document and the user profile; and

instructions for ranking at least the plurality of the search result documents according to their personalized scores.

44. The computer program product of claim 43, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have been viewed by the user.

45. The computer program product of claim 43, wherein the set of documents include a plurality of documents that have been identified by search results from the search engine and that have not been viewed by the user.

46. The computer program product of claim 43, further including:

instructions for updating a term-based profile by identifying a set of terms from a document in the set of documents, and adding information about the identified set of terms to the term-based profile; and

instructions for updating a category-based profile by classifying the document into a plurality of categories, and adding information about the plurality of categories to the category-based profile.

47. The computer program product of claim 46, further including:

instructions for updating a link-based profile by analyzing links in the document, and adding information derived from the analyzed links to the link-based profile.

48. The computer program product of claim 46, wherein the link-based profile includes information about URLs or portions of URLs.

49. The computer program product of claim 48, wherein the link-based profile comprises:
a plurality of URLs and a weight associated with each URL, wherein the weight is based on one or more factors selected from the group consisting of frequency with which the user visits the URL, time the user has spent viewing a document associated with the URL and quantity of the user's scrolling activity at the document; and

a plurality of hosts and a weight associated with each host, wherein the weight is based on frequency of the user's visits to the host.

50. The computer program product of claim 49, wherein the URLs further include URLs that have not been visited by the user, but are related to the URLs that have been visited by the user and the weight of an unvisited URL depends on its distance to at least one related URLs that have been visited.

51. The computer program product of claim 46, wherein a term in the term-based profile is an expression comprising at least one word and a weight.

52. The computer program product of claim 48, wherein the weight is a weight associated with occurrences of the term in the set of documents.

53. The computer program product of claim 48, wherein the weight of a term depends at least partially on the term's term frequency and inverse document frequency in said set of documents.

54. The computer program product of claim 46, wherein a category in the category-based profile characterizes at least one aspect of documents in the category and the category is associated with a weight indicative of the category's importance relative to other categories.

55. The computer program product of claim 54, wherein the at least one aspect of the documents in the category is selected from the group consisting of: document format, document type, document topic and document origin.

56. The computer program product of claim 43, wherein the user profile is stored on a server of the search engine.

57. The computer program product of claim 43, wherein the user profile is stored on a client associated with the user.

58. The computer program product of claim 43, wherein the user is a group of users.